New Technology Chips Away at Language Barriers

By WILLIAM MATTHEWS

In biblical lore, Babylon, the ancient city 55 miles south of Baghdad, was the place where an angry God divided his sinful people, making them speak different languages and then scattering them.

Today, it's where technology is helping to overcome millennia-old language barriers. U.S. soldiers in Iraq are able to communicate with Iraqis to an ever greater degree through new electronic devices.

Conversations may be as simple as the one-way, English-to-Arabic translation of basic commands at checkpoints, such as "stop" and "get out of the car." Or they may be more complex two-way conversations.

Since the war began in 2003, the Pentagon and private companies have poured millions of dollars into speech translation research, and the results are paying off in steadily improving translation gear. The U.S. Defense Advanced Research Projects Agency has been investing in the technology at a rate of $15 million to $20 million a year.

What translation systems can do now "is two or three times more" than they could do in the early years of the war, said Craig Schlenoff, an electronics engineer with the Intelligent Systems Division of the National Institute of Standards and Technology (NIST).

The most advanced systems are designed to translate questions from English to Iraqi Arabic and answers from Arabic back to English.

They're "very good at what they're trained to do," said Schlenoff, who heads a NIST program that evaluates translation hardware and software for the military. "If you feed in a lot of data for a domain, most of the systems will do a good job."

A "domain" is a set of words, phrases, questions and answers likely to be used under certain circumstances. One domain would be words and phrases related to operating a checkpoint. Others involve medical emergencies, house searches, asking directions, conducting maritime interdiction operations, crowd control and so on.

The best translation devices are well enough designed that "they know the questions, they know the expected answers and they can guess pretty well" what an answer is, despite awkward grammatical construction or an unusual speaking accent, Schlenoff said.

"Are they perfect? No. Are they good? Yes," he said.

For the foreseeable future, a translating machine is never going to be as good as a human translator, who "may turn to you and say: This guy's lying," Schlenoff said. Body language and other visual and audible clues may make it obvious that the answers being given are untrue, he said. So far, no translating machine is able to do that.

VCOM3D PHOTO

Hands-Free Translating: At the U.S. Army's Joint Readiness Training Center, Fort Polk, La., a soldier demonstrates how the Vcommunicator iPod-sized translator device fits on his wrist but doesn't interfere with firing his weapon.

But if no human translator is available — a common circumstance in Iraq — the machine "is a huge step forward," Schlenoff said.

The Simple: Vcommunicator

Among the newest of mechanical translators: the iPod. A small Florida firm, Vcom3D, developed software that turns an Apple iPod into a translator that is able to make relatively simple statements and ask yes-or-no questions in Arabic, Kurdish, Pashto or Dari.

"You pick the language you want, you pick the mission and then you pick the phrase you want to speak," said Ernie Bright, product manager for Vcom3D's Vcommunicator.

For manning a checkpoint, the user selects prerecorded phrases such as, "We need to search your vehicle," and "Stop your vehicle," and these are spoken by the iPod.

To aid in understanding, the command also appears in Arabic script on the iPod's screen. And if needed, an animated character can act out gestures and facial expressions on the iPod screen, Bright said.

The Vcommunicator is customizable. New phrases can be added, and so can photos, maps and videos to help soldiers searching for individuals and finding locations.

Besides serving as a translator, the Vcommunicator is also intended to teach foreign phrases to the soldiers who use it, Bright said.

Vcommunicators, which cost about $2,000 apiece, were kept simple so they would be easy to use with little training, Bright said.

The Complex: MASTOR

At the other end of the complexity scale is two-way translation software developed by IBM. It's called MASTOR, for Multilingual Automatic Speech Translator.

The intent here was to create software that can translate English into another language, and then translate responses in that language back into English.

What appears to be a two-step process is actually much more complex. Three technologies are needed for the process to work — speech recognition, text translation and text to speech, said David Nahamoo, chief of IBM's human language technology group.

First, the speech recognition software turns the English speech into English text. Next the text is translated into the Arabic text, which then is turned into spoken Arabic.

That process is substantially complicated by the nature of speech.

In speech, more often than in writing, "you have to deal with nongrammatical sentences, repeats, hesitation, corrections, lots of uhs and ums, lack of punctuation to show where one sentence ends and another starts," Nahamoo said.

The meaning of words creates another set of problems. The English word "may," for example, can mean a month, can be a request for permission and can be a statement of probability. The translating software has to figure out which way it is being used in order to use the right word in Arabic.

IBM's uses a "very data-driven statistical technique" that relies on word counts and the sequencing of words in the two languages to determine meaning, Nahamoo said.

Where a word appears in a sentence and what word appears next to it helps the software hypothesize the sentence's meaning, he said. The software scores possible interpretations and then chooses the one with the highest probability. This happens more or less instantly.

When the English is ready to be spoken in Arabic, English text appears on a screen so that the speaker can tell if the translation from speech to text is accurate. If it is not, a second or third option can be selected, translated into Arabic and then turned into Arabic speech.

Like other translators, IBM's is built around domains.

The ultimate goal is to create a translator that can handle unlimited vocabulary, Nahamoo said. That's still a long way off. "But when you draw a domain boundary around it, we have made enough progress to deliver good accuracy" in translating speech from one language to another and back.

"Every couple of months, we make improvements in the technology," he said. "If you compare what we have today with two years ago, the difference is night and day. It's a lot more accurate, it has more vocabulary and faster response."

The Original: Phraselator

The Phraselator was one of the earliest mechanical translators adopted by the U.S. military.

The originals, circa 2004, used English speech recognition to select and play prerecorded phrases in one of about 30 languages. The device cannot understand spoken responses, so phrases are limited to instructions that require an action in response, such as "Get out of the car."

A phrase-based two-way translator is coming out next fall and a "directed dialog" version is due in 2009, said John Hall, president of Voxtec International, the Phraselator's maker.

These two-way translators, like others, are built around domains. Questions or statements are programmed into the translator, and so are expected answers.

So to the question "What is your age," the translator will expect an answer that includes numbers. It would recognize the answer "36" or "I am 36 years old," but would not understand the answer, "I am two years older than my brother."

Directed dialog takes this a step further by permitting simple but nearly normal conversation within a restricted domain, similar to the technology used by automated telephone answering systems.

Voxtec's new translators will have more memory and greater computing power to handle larger domains, Hall said.

The Specialist: VRT

Integrated Wave Technologies (IWT) President Tim McCune isn't convinced all that is necessary. IWT produces the Voice Response Translator, a small, one-way limited-response translator that McCune credits with helping U.S. soldiers capture a car-bomb maker, conduct house searches and arrest insurgents.

"It does a limited thing well," McCune said.

About 5,000 of the devices are in use in Iraq. In Afghanistan, they are also used to communicate with Afghan National Army troops during training.

Rather than scan screens to find appropriate phrases for the translator to speak, the VRT is trained to recognize its user's voice and translate selected phrases spoken into the microphone on a headset.

"It's truly eyes-free and hands-free," McCune said — enabling soldiers to communicate without taking their eyes off those they're speaking to or their hands off their weapons.

The VRT has a 65-hour battery life, weighs 11 ounces and works well amid ambient noise, McCune said. Upcoming versions will be smaller and better, he said. Among the planned improvements: limited capability for the VRT to translate responses back to English.

This year, IWT is testing the capability to recognize about 5,000 Arabic nouns and adjectives, he said.

The Compact: IraqComm

SRI International is trying to combine the best attributes of today's translators by offering complex two-way, speech-to-speech translation packed into ever smaller, rugged computers.

As its name suggests, SRI's IraqComm translator was developed specifically for the Iraq war. Like IBM's, it links speech-to-text technology with machine translation and text-to-speech synthesis.

The company's latest translation package runs on a laptop that's about the same size as the Phraselator, said Kristin Precoda, director of SRI's speech technology and research laboratory.

IraqComm offers users a range of configurations, including hands-free, eyes-free operation for basic communication, she said.

Like the others, IraqComm translates within predetermined domains. "It's pretty conversational within the domains it's trained for," Precoda said.

In tests conducted by NIST, she said, an English speaker and an Arabic speaker were given a limited amount of time to assign each other tasks through the translator. NIST judged the translators by how well each person carried out his task.

"They got it more than 90 percent of the time," Precoda said. ■

E-mail: bmatthews@defensenews.com.